

009873221

WPI Acc No: 94-153134/\*199419\*

**Speech discrimination in presence of ambient noise for voice encoder -  
analysing signal comprising sum of aperiodic and predetermined number of  
periodic excitations, by calculating global auto-correlation and partial  
sums of auto-correlation**

Patent Assignee: THOMSON CSF (CSFC )

Inventor: LAURENT P A

Number of Countries: 004 Number of Patents: 002

Patent Family:

| Patent No  | Kind | Date     | Applicat No | Kind | Date     | Main IPC    | Week     |
|------------|------|----------|-------------|------|----------|-------------|----------|
| EP 596785  | A1   | 19940511 | EP 93402670 | A    | 19931029 | G10L-009/14 | 199419 B |
| FR 2697937 | A1   | 19940513 | FR 9213397  | A    | 19921106 | G10L-007/02 | 199422   |

Priority Applications (No Type Date): FR 9213397 A 19921106

Cited Patents: 5.Jnl.Ref

Patent Details:

| Patent | Kind | Lan | Pg | Filing Notes | Application | Patent |
|--------|------|-----|----|--------------|-------------|--------|
|--------|------|-----|----|--------------|-------------|--------|

|           |    |   |    |  |  |  |
|-----------|----|---|----|--|--|--|
| EP 596785 | A1 | F | 13 |  |  |  |
|-----------|----|---|----|--|--|--|

Designated States (Regional): DE GB NL

Abstract (Basic): EP 596785 A

The speech discrimination method involves analysing a signal (Sn) consisting of the sum of a given number (K) of periodic excitations and an aperiodic excitation. It includes calculating (1) the global autocorrelation (rm) of the signal (Sn), and calculating (2) the partial sums (tm) of the autocorrelation with aperiodic short term.

A counter (3) is initialised such that the counter value (k) does not achieve the maximum value (K) corresp. to the maximum number of periodic excitations. For each incrementation of the counter, having corrected the calculation of partial sums, several parameters are determined. These include the pitch (Mk), the gain (beta-k) and the slope of the gain (alpha-k) for each periodic excitation. The level of the aperiodic excitation (beta-o) is also determined.

ADVANTAGE - Allows discrimination of speech in presence of background noise for low data rate voice encoder.

Dwg.1/4

Derwent Class: P86; W04

International Patent Class (Main): G10L-007/02; G10L-009/14

International Patent Class (Additional): G10L-009/08

**This Page Blank (uspto)**



12 **DEMANDE DE BREVET EUROPEEN**

21 Numéro de dépôt : **93402670.9**

61 Int. Cl.<sup>6</sup> : **G10L 9/14, G10L 9/08**

22 Date de dépôt : **29.10.93**

30 Priorité : **06.11.92 FR 9213397**

43 Date de publication de la demande :  
**11.05.94 Bulletin 94/19**

84 Etats contractants désignés :  
**DE GB NL**

71 Demandeur : **THOMSON-CSF**  
**173, Boulevard Haussmann**  
**F-75008 Paris (FR)**

72 Inventeur : **Laurent, Pierre André,**  
**Thomson-CSF, SCPI**  
**BP 329**  
**F-92402 Courbevoie Cédex (FR)**

74 Mandataire : **Lincot, Georges et al**  
**THOMSON-CSF, SCPI, B.P. 329, 50, rue**  
**Jean-Pierre Timbaud**  
**F-92402 Courbevoie Cédex (FR)**

54 **Procédé de discrimination de la parole en présence de bruits ambiants et vocodeur à faible débit pour la mise en oeuvre du procédé.**

57 Le procédé de discrimination de la parole en présence de bruits ambiants consiste pour analyser un signal  $S_n$  composé de la somme d'un nombre  $K$  déterminé d'excitations périodiques et d'une excitation aperiodique, à calculer (1) l'autocorrélation globale  $r_m$  du signal  $S_n$ , à calculer (2) les sommes partielles  $t_m$  de l'autocorrélation à court terme aperiodique; à calculer (1) l'autocorrélation globale  $r_m$  du signal  $S_n$ , à calculer (2) les sommes partielles  $t_m$  de l'autocorrélation à court terme  $s_m$  corrélée avec l'autocorrélation globale  $r_m$ , à initialiser (3) un compteur  $k$  et tant que le compteur  $k$  n'atteint pas la valeur maximale  $K$  correspondant au nombre maximal d'excitations périodiques (6), pour chaque incrémentation (4) du compteur  $k$ , après avoir corrigé (7) le calcul des sommes partielles  $t_m$ , à calculer (5) les valeurs du pitch  $M_k$ , du gain  $\beta_k$  et de la pente du gain  $\alpha_k$  de chaque excitation périodique, et à déterminer le niveau de l'excitation aperiodique  $\beta_0$  en fonction de l'autocorrélation globale  $r_m$  en début d'analyse et de l'autocorrélation globale  $r_m$  en fin d'analyse.

Application : vocodeurs faible débit.

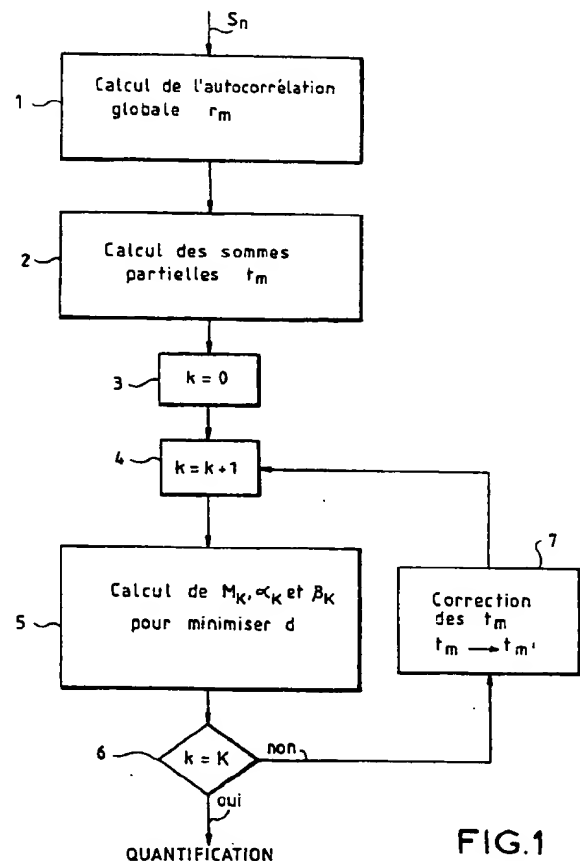


FIG.1

La présente invention concerne un procédé de discrimination de la parole en présence de bruits ambiants et un vocodeur à faible débit pour la mise en oeuvre du procédé.

La qualité des vocodeurs de type LPC10 (prédiction linéaire d'ordre 10) est souvent jugée insuffisante, notamment en termes d'agrément d'écoute, de fidélité au locuteur, et de résistance aux bruits ambiants notamment aux bruits ambiants structurés de nature périodique ou quasi-périodique.

Le principe même de ce type de vocodeur conduit à une fidélité limitée car il part de l'hypothèse que le signal à coder est exclusivement de la parole et, qui plus est, que cette parole peut être représentée sous une forme simple telle que du bruit filtré, ou une excitation périodique elle aussi filtrée.

Si le signal à coder ne répond pas à cette définition, par exemple un signal semi-périodique, ou un mélange de plusieurs signaux, la qualité de reproduction est médiocre.

Pour améliorer la qualité de reproduction de la parole, différentes méthodes connues, consistent à utiliser, par exemple, un vocodeur haut débit de 4800 bits/seconde et à réduire ce débit à 2400 bits/seconde. Aucune hypothèse n'est faite sur la nature du signal à coder, le but de ces méthodes consistant uniquement à reproduire le plus fidèlement possible la forme d'onde du signal d'entrée.

Parmi ces méthodes, on peut citer la méthode dite des ondelettes qui est une représentation du signal par une combinaison de formes d'onde bien localisées en temps et en fréquence, l'analyse harmonique qui est une représentation du signal par une combinaison de sinusoides harmoniques les unes des autres et, le CELP abréviation anglo-saxonne pour Code Excited Linear Prediction ou les formes d'ondes utilisées à l'entrée du filtre de synthèse sont pré-définies, et stockées dans un "dictionnaire".

L'expérience montre que, si l'on cherche à réduire par trop le débit des vocodeurs qui traitent la forme d'onde du signal, la qualité de reproduction s'en ressent beaucoup. Les dégradations de la qualité se traduisent, par exemple, par une certaine raucité de la parole synthétique, et/ou un bruit de fond de coloration variable au cours du temps, et/ou une grande difficulté à reproduire les sons bruités ou au contraire périodiques.

Le but de l'invention est de pallier les inconvénients précités.

A cet effet, l'invention a pour objet un procédé de discrimination de la parole en présence de bruits ambiants pour vocodeur à faible débit du type comportant une excitation périodique, une excitation apériodique et un filtre d'analyse d'ordre  $P$ , caractérisé en ce qu'il consiste pour analyser un signal  $S_n$  composé de la somme d'un nombre  $K$  déterminé d'excitations périodiques et d'une excitation apériodique, à calculer l'autocorrélation globale  $r_m$  du signal  $S_n$ , à calculer les sommes partielles  $t_m$  de l'autocorrélation à court terme  $s_m$  corrélée avec l'autocorrélation globale  $r_m$ , à initialiser un compteur  $k$  et tant que le compteur  $k$  n'atteint pas la valeur maximale  $K$  correspondant au nombre maximal d'excitations périodiques, pour chaque incrémentation du compteur  $k$ , après avoir corrigé le calcul des sommes partielles  $t_m$ , à calculer les valeurs du pitch  $M_k$ , du gain  $\beta_k$  et de la pente du gain  $\alpha_k$  de chaque excitation périodique, et à déterminer le niveau de l'excitation apériodique  $\beta_0$  en fonction de l'autocorrélation globale  $r_m$  en début d'analyse et de l'autocorrélation  $r_m$  en fin d'analyse.

Le procédé selon l'invention a pour principal avantage qu'il permet de reproduire une parole de meilleure qualité qu'avec un vocodeur standard à 2400 bits/seconde et de mieux résister aux bruits ambiants et notamment aux bruits ambiants structurés. Il a également pour avantage d'utiliser un algorithme de complexité raisonnable limitant ainsi la charge de calcul.

D'autres caractéristiques et avantages de l'invention apparaîtront dans la description qui suit faite en regard des dessins annexés qui représentent :

- la figure 1, les différentes étapes d'un premier mode de réalisation du procédé selon l'invention mises sous la forme d'un organigramme ;
- la figure 2, les diagrammes représentant l'évolution de l'autocorrélation ;
- la figure 3, les différentes étapes d'un second mode de réalisation du procédé mises sous la forme d'un organigramme ;
- la figure 4, un mode de réalisation d'un vocodeur pour la mise en oeuvre du procédé selon l'invention.

Le procédé selon l'invention repose sur le principe qu'il n'est pas utile de reproduire la forme d'onde du signal d'entrée et qu'il faut plutôt reproduire du mieux possible l'impression auditive qu'aurait produit le signal original, ce qui n'est pas obligatoirement la même chose: un vocodeur standard à 2400 bits/seconde pouvant restituer un signal de parole d'excellente qualité avec certains locuteurs et dans de bonnes conditions de prise de son, bien que la forme d'onde produite à la synthèse n'a que peu de choses à voir avec la forme d'onde originale.

Le débit alloué au filtre de prédiction n'étant pas suffisant pour représenter le signal avec une fidélité suffisante, il faut le modéliser. Pour cela, la parole synthétique est considérée donner une impression acoustique voisine de celle procurée par le signal de parole original. La parole synthétique ainsi considérée est constituée de la superposition de formes d'ondes particulièrement simples qui peuvent être définies avec un faible débit binaire. Pour définir ces formes d'ondes, le vocodeur standard à 2400 bits/seconde est supposé donner une qualité satisfaisante dans des cas simples, par exemple, dans des cas où le signal à coder peut être représenté

comme la superposition d'un bruit de fond continu, et d'un ou plusieurs signaux périodiques ou quasi-périodiques ; la même hypothèse est faite dans les vocodeurs harmoniques.

D'autre part, dans le cas où apparaît un transitoire brutal tel qu'un bruit extérieur, la fidélité de reproduction du signal d'entrée n'est vraisemblablement pas indispensable. Le transitoire est alors considéré comme un son parasite et il est éliminé par des systèmes d'antiparasitage connus améliorant ainsi la qualité de restitution du signal d'entrée.

De plus, le débit accordé au filtre de prédiction d'un vocodeur standard peut être réduit suivant des techniques connues, utilisées dans les vocodeurs à 800 bits/seconde, permettant ainsi de libérer des bits alloués au filtre de prédiction trop richement décrit.

Les bits ainsi récupérés sont utilisés pour définir les K excitations périodiques ayant chacune une période ou "pitch" déterminée et un gain modulable dans le temps.

Un premier mode de réalisation du procédé selon l'invention consiste à déterminer le signal d'excitation représentatif de la parole extraite des bruits ambiants, en donnant la période du "pitch" et le niveau du signal à partir de la somme des K signaux périodiques et d'un signal apériodique.

Dans un vocodeur standard de 2400 bits/seconde, il est obligatoire de déterminer un seul pitch sans erreur et un indicateur de voisement également sans erreur. Avantagusement, selon l'invention, si le premier générateur périodique, matérialisant l'excitation, ne dispose pas du "vrai" pitch, c'est-à-dire du pitch du signal de parole à extraire, il reste encore K-1 générateurs pour le trouver. De même, comme il n'y a pas de décision de voisement mais plutôt une répartition des gains entre une source de bruit et K sources périodiques, il n'y a pas de risque d'erreur de voisement.

Le procédé selon l'invention ne s'intéresse pas à la forme d'onde vraie, ou à un résiduel, mais à sa composition en termes de périodes ou "pitch", niveaux relatifs, et proportion de bruit. La détermination de l'excitation se fait donc à partir d'un signal où les informations de phase n'apparaissent pas. Le procédé est basé sur un calcul d'autocorrélation dont le résultat donne une grandeur représentative de ce signal moyennant certaines précautions pour détecter les composantes périodiques et les variations de gain.

La figure 1 représente l'organigramme des étapes du procédé selon l'invention.

La première étape 1 calcule l'autocorrélation globale  $r_m$  du signal constitué de la somme des K + 1 signaux.

Si P représente l'ordre du filtre de prédiction à court terme, et  $M_{\max}$  la valeur maximale de pitch considérée en nombre d'échantillons, l'autocorrélation globale  $r_m$  est calculée pour  $m = -P$  à  $m = M_{\max} + P$ , suivant la formule :

$$(1) \quad r_m = \frac{\sum_{n=0}^{N(m)-1} S_n S_{n-m}}{\sum_{n=0}^{N(m)-1} S_n^2}$$

Dans l'expression (1)  $S_n$  et  $S_{n-m}$  sont des amplitudes d'échantillons de signal et  $N(m)$  désigne un nombre d'échantillons multiple de m, le plus grand qui soit inférieur ou égal à une valeur  $N_{\max}$ . Cette disposition permet d'améliorer la détection ultérieure des périodicités.

La valeur choisie pour le dénominateur permet de détecter la variation du niveau de signal, donc d'améliorer la fidélité. Par exemple, pour un niveau de signal multiplié par g sur une durée de m échantillons, le signal ayant une période m, on obtient  $r_m = g \cdot r_0$ ,  $r_{2m} = g^2 \cdot r_0$ , etc...

Les valeurs de l'autocorrélation globale  $r_m$  sont ensuite analysées en faisant comme première hypothèse simplificatrice que l'autocorrélation à court terme du signal est nulle pour un écart m supérieur à P en valeur absolue : ceci devient presque vrai si l'on utilise non pas le signal original mais plutôt le résiduel à la sortie d'un prédicteur auto-adaptatif d'ordre suffisant et ayant une constante de temps d'adaptation bien choisie.

Par exemple, il peut être avantageux de remplacer  $S_n$  par un résiduel  $S'_n$  donné par les équations suivantes :

$$(2) \quad S'_n = S_n - \sum_{i=1}^Q A_i^n S_{n-i}$$

avec  $A_i^{n+1} = A_i^n + \varepsilon \text{ signe}(S'_n S_{n-i})$ ,  $i=1 \dots Q$

Dans une deuxième hypothèse, la valeur minimale  $M_{\min}$  du pitch est choisie supérieure à 2P.

En tenant compte des deux hypothèses précédentes, l'autocorrélation globale  $r_m$  est ensuite recomposée à partir de la somme des excitations périodiques  $M_1, M_2, \dots, M_K$  et des valeurs de l'autocorrélation à court terme  $r_p, \dots, r_p$  dupliquées aux positions  $0, M_1, 2M_1, \dots, 0, M_2, 2M_2, \dots, 0, M_K, 2M_K, \dots$ , les répliques suivant l'évolution du niveau des composantes du signal.

La figure 2 représente, de haut en bas, l'évolution de l'autocorrélation  $r_m$  en fonction de  $m$ , pour  $m = -P$  à  $m = M_{\max} + P$ . Sur cette figure, la contribution du bruit n'est pas représentée. Le premier diagramme représente l'autocorrélation à court terme, les deux diagrammes suivants, la contribution des signaux périodiques  $M_1$  à  $M_K$ , et le dernier diagramme l'autocorrélation globale  $r_m$  obtenue à partir des valeurs de l'autocorrélation à court terme  $r_p, \dots, r_p$ , convoluées avec  $K$  trains d'impulsions.

Afin de tenir compte de l'évolution du niveau du signal, le train d'impulsions relatif à la kème excitation périodique est défini par la formule suivante:

$$I_{k,n} = \beta_k \sum_{i=0}^{\text{INT}\left(\frac{M_{\max}}{M_k}\right)} (1 + i\alpha_k) \delta(n - iM_k) \quad (3)$$

Le coefficient  $\beta_k$  représente un gain, le coefficient  $\alpha_k$  une variation de gain, ou pente de gain qui doit être linéaire pour pouvoir poursuivre les calculs et  $\text{INT}(M_{\max}/M_k)$  est une fonction qui ne retient que la partie entière du rapport  $M_{\max}/M_k$  soit le rapport entre la valeur du pitch maximum et la valeur du pitch de la kème excitation périodique.

Compte tenu des deux hypothèses précédentes concernant les valeurs respectives de  $P$ , ordre de prédiction, et  $M_{\min}$ , valeur minimale recherchée pour le pitch  $M_k$ , l'étape 2 sur la figure 1 consiste à calculer les sommes partielles  $t_m$  qui correspondent en fait au calcul de l'autocorrélation de l'autocorrélation globale  $r_m$  limitée à sa valeur à court terme. Le calcul est donné par la formule suivant :

$$t_m = \sum_{i=-P}^P s_i r_{i+m} \quad \text{avec } m = 0 \dots M_{\max}$$

et

$$t_0 = \sum_{i=-P}^P s_i^2 = \sum_{i=-P}^P r_i^2 \quad (4)$$

$s_p, \dots, s_p$  sont les valeurs de l'autocorrélation  $r_p, \dots, r_p$  limitée à sa valeur à court terme.

Pour des raisons de complexité de calcul, un algorithme itératif de recherche sub-optimal pour trouver les  $K$  valeurs de  $M_k$ ,  $\alpha_k$  et  $\beta_k$  correspondant respectivement à la période du pitch, la pente du gain et le gain de la kème excitation est mis en oeuvre à l'étape 5. Il consiste à calculer les valeurs de  $M_k$ ,  $\alpha_k$  et  $\beta_k$  qui minimisent la quantité  $d$  suivante, par exemple, par une méthode des moindres carrés :

45

$$5) \quad d = \sum_n \left( r_n - \beta_k \sum_i (1 + i\alpha_k) s_{n - iM_k} \right)^2$$

50

Les étapes 3 et 4 correspondent respectivement à l'initialisation d'un compteur  $k$  et à l'incrément de ce compteur  $k$  tant que la valeur du compteur  $k$  n'a pas atteint la valeur  $K$ . Ce test est effectué par l'étape 6 du procédé selon l'invention.

En notant  $R$  le vecteur des autocorrélations  $r_m$  et  $Sl_k$  celui des autocorrélations à court terme  $s_m$  convoluées avec le kème train d'impulsions, avec un gain égal à 1, la quantité  $d$ , représentant la distance entre le vecteur  $R$  et le vecteur  $Sl_k$  affecté d'un gain  $\beta_k$ , est obtenue par la formule suivante :

$$(6) \quad d = |R - \beta_k Sl_k|^2$$

avec

$$R = \{r - p, \dots, R_{M_{\max} + p}\}$$

$$Sl_k = \{sl_k - p, \dots, sl_{k, M_{\max} + p}\}$$

5

$$si_{k,n} = \sum_i (1 + i \alpha_k) s_{n-i} M_k$$

10

et

La valeur de  $\beta_k$  qui minimise cette expression (6) est obtenue par la formule suivante :

$$(7) \quad \beta_k = \frac{R \cdot Sl_k}{|Sl_k|^2}$$

15

Dans ces conditions, la quantité  $d$  devient  $d_{\min}$  telle que :

$$d = d_{\min} = |R|^2 - \frac{(R \cdot Sl_k)^2}{|Sl_k|^2} \quad (8)$$

En reprenant le calcul des sommes partielles donné par la relation (4), l'expression  $d_{\min}$  donnée par la relation (8) devient alors :

20

$$d_{\min} = |R|^2 - \frac{\left( \sum_i (1 + i \alpha_k) t_i M_k \right)^2}{t_0 \sum_i (1 + i \alpha_k)^2} \quad (9)$$

25

Il ne reste plus qu'à minimiser cette expression en fonction de  $\alpha_k$ , ce qui donne, tous calculs effectués, la valeur suivante pour  $\alpha_k$  :

30

$$(10) \quad \alpha_k = \frac{S_0 T_1 - S_1 T_0}{S_2 T_0 - S_1 T_1}$$

avec

35

$$S_k = \sum_{i=0}^{\text{INT}\left(\frac{M_{\max}}{M_k}\right)} i^k$$

40

et

45

$$T_k = \sum_{i=0}^{\text{INT}\left(\frac{M_{\max}}{M_k}\right)} i^k t_i M_k$$

50

La valeur de  $d_{\min}$  ainsi obtenue est alors la suivante :

$$(11) \quad d_{\min} = |R|^2 - \frac{S_2 T_0^2 - 2 S_1 T_0 T_1 + S_0 T_1^2}{t_0 (S_0 S_2 - S_1^2)}$$

55

La recherche d'une excitation donnée consiste à trouver la valeur  $M_k$  qui minimise cette quantité, sachant que  $R$  ne varie pas durant la recherche et que les quantités  $S_0$ ,  $S_1$  et  $S_2$  sont facilement calculables pour une valeur de  $M_k$  donnée.

Une fois que la valeur de  $M_k$  a été déterminée, l'équation (10) donne la valeur de  $\alpha_k$ , et la valeur de  $\beta_k$  est donnée par modification de l'équation (7):

$$(12) \quad \beta_k = \frac{S_2 T_0 - S_1 T_1}{t_0(S_0 S_2 - S_1^2)}$$

Lors de l'itération suivante, c'est-à-dire au passage de la valeur  $k$  à la valeur  $k + 1$  du compteur  $k$  correspondant à l'étape 4 sur la figure 1, le vecteur  $R$  des autocorrélations  $r_m$  n'est que partiellement modélisé par le vecteur  $S_k$  multiplié par le gain  $\beta_k$ . Il convient donc de remplacer les autocorrélations  $r_m$  par leurs valeurs modifiées  $r'_m$  par soustraction des quantités  $\beta_k s_{i,k,m}$  suivant l'équation suivante :

$$(13) \quad r_m \rightarrow r'_m = r_m - \beta_k \sum_{i=0}^{\text{INT}\left(\frac{M_{\max}}{M_k}\right)} (1 + i \alpha_k) s_{m - i M_k}$$

Cependant, comme les valeurs de l'autocorrélation globale  $r_m$  n'interviennent pas directement dans le calcul de  $\alpha_k$  et  $\beta_k$ , le procédé ne modifie que les sommes partielles  $t_m$  qui sont seules utilisées. Cette correction correspondant à l'étape 7 consiste à soustraire des sommes partielles  $t_m$  les valeurs de l'autocorrélation  $c_m$  des échantillons  $s_m$  de l'autocorrélation à court terme et à remplacer les sommes partielles  $t_m$  par leurs valeurs modifiées  $t'_m$ . Ceci est réalisé suivant la relation suivante :

$$(14) \quad t_m \rightarrow t'_m = t_m - \beta_k \sum_{i=0}^{\text{INT}\left(\frac{M_{\max}}{M_k}\right)} (1 + i \alpha_k) c_{m - i M_k}$$

avec

$$c_q = \sum_{n=\text{MAX}(-P, -P-q)}^{n=\text{MIN}(P, P-q)} S_n S_{n+q}$$

Les coefficients  $c_m$  ne sont calculés qu'une fois, puisque  $s_m$  ne change pas au cours des itérations.

Une fois les  $K$  excitations périodiques définies, le niveau de l'excitation aperiodique à utiliser, se déduit de l'autocorrélation  $r_m$ . En effet, si le signal répondait exactement à un modèle à  $K$  excitations périodiques, la valeur de l'autocorrélation  $r_m$  ou des sommes  $t_m$  serait nulle après la dernière correction suivant les équations (13) et (14). En pratique, le niveau  $\beta_0$  de bruit à générer peut être représenté, compte tenu des approximations faites, par une fonction non décroissante du rapport entre les valeurs des coefficients  $r_m$  après définition de la  $K$ ème excitation, et leurs valeurs initiales. Cette fonction peut être par exemple une relation empirique du type :

$$(15) \quad \beta_0 = F \left( \frac{\sum_{m=0}^{M_{\max}} |r_{m \text{ final}}|^\gamma}{\sum_{m=0}^{M_{\max}} |r_{m \text{ initial}}|^\gamma} \right),$$

avec  $\gamma = 1$  ou  $2$

Bien entendu, le procédé décrit ci-dessus peut être amélioré sur divers points qui sont abordés brièvement ci-après. Un nouvel organigramme des étapes d'un deuxième mode de réalisation du procédé selon l'invention est représenté sur la figure 3. Sur cette figure les étapes homologues à celles de la figure 1 sont désignées par les mêmes repères.

L'étape 8 du procédé consiste en un prétraitement du signal d'entrée. Ce prétraitement transforme, par exemple, le signal d'entrée  $S_m$  brut en un signal dont l'autocorrélation se rapproche d'une impulsion de dirac, donc d'un signal dont le spectre est aplati, par exemple, par un filtre prédictif auto-adaptatif. Ce prétraitement



réalise ainsi un blanchissement du spectre avant analyse.

D'autres prétraitements, tels que par exemple, l'élimination de la composante continue et des très basses fréquences du signal d'entrée, un contrôle automatique de gain, et une pré-accentuation, sont également possibles.

5 A partir du signal ainsi traité par l'étape 8, le calcul de l'autocorrélation globale  $r_m$  est effectué selon l'étape 1 décrite précédemment. Une deuxième étape 9 supplémentaire consiste à pondérer l'autocorrélation venant d'être calculée par une fenêtre de pondération simple pouvant être représenté par exemple par une enveloppe non croissante en fonction du temps et dont la largeur est choisie plus large que l'intervalle maximal d'analyse. Le but de cette fenêtre de pondération vise plus à stabiliser le signal plutôt qu'à le mettre en forme en évitant  
10 des discontinuités dans la suite des calculs dues au nombre variable de répliques de l'autocorrélation à court terme que peuvent comporter les vecteurs  $Sl_k$ . Par exemple, si  $M_{\max} = 160$ , il y aura trois répliques pour un pitch de 80 et seulement deux répliques pour un pitch de 81. Cette étape fait donc apparaître progressivement les répliques de l'autocorrélation dont les amplitudes suivent une fonction décroissante.

Les étapes 2, 3 et 4 se retrouvent à l'identique de la figure 1, et l'étape 5 est pratiquement identique moyennant une limitation sur les valeurs de  $\beta_k$  et  $\alpha_k$  :

En effet, dans le cas d'un signal complètement non voisé, par exemple un signal de parole sans voyelle, le procédé selon l'invention est capable de déterminer les K pitches recherchés. La seule différence avec un son partiellement ou totalement voisé réside dans la valeur des coefficients  $\beta$  et  $\alpha$ . Pour éviter cette situation, le calcul selon l'étape 5 ne retient que les excitations périodiques pour lesquelles les coefficients  $\beta$  et  $\alpha$  sont  
20 compris dans des plages de valeurs restreintes : par exemple, des valeurs positives inférieures à 1 pour  $\beta \geq 0,3$  et  $\beta \leq 1$ , et des valeurs proches de 1 pour  $\alpha$ ,  $|\alpha| = 0,2$ . La limitation des valeurs de  $\alpha$  permet également d'éviter des impulsions de diracs négatives représentatives de l'autocorrélation. Le coefficient  $\alpha$  peut répondre par exemple à la relation suivante :

$$(16) \quad |\alpha| \leq \frac{1}{\text{INT}(M_{\max}/M_k)}$$

25 L'étape 10 du procédé, consiste en un test supplémentaire sur la valeur du compteur k à l'issue du calcul des coefficients  $M_k$ ,  $\alpha_k$  et  $\beta_k$  effectué par l'étape 5 du procédé.

Ce test prend en compte l'éventualité que la première excitation analysée,  $k < 2$ , par le procédé soit le signal de parole recherché. Dans ce cas, l'étape 10 est rebouclée sur l'incréméntation du compteur k représentée  
30 par l'étape 4.

Dans l'autre cas,  $k \geq 2$ , l'étape 11 du procédé recalcule les coefficients calculés par l'étape 5: l'algorithme utilisé par le procédé selon l'invention est sub-optimal, c'est-à-dire qu'il recherche les K excitations périodiques les unes après les autres, alors qu'en toute rigueur il devrait les rechercher toutes à la fois. De plus, les vecteurs  $Sl_k$  ne sont pas orthogonaux, ils partagent tous les autocorrélations  $r_p$  à  $r_p$  créant des interférences entre les  
35 diverses autocorrélations. Afin d'optimiser l'algorithme, l'étape 11 recalcule les coefficients  $\alpha_1, \alpha_2, \dots, \alpha_{k-1}$  et  $\beta_1, \beta_2, \dots, \beta_{k-1}$  en plus de  $\alpha_k$  et  $\beta_k$  à la même itération, en conservant les valeurs de pitches  $M_k$  précédemment calculés; ce qui revient à une résolution d'un système de K équations linéaires.

Les étapes 6 et 7 suivantes correspondent respectivement à celles du procédé de la figure 1.

Une dernière correction est apportée au premier mode de réalisation du procédé selon l'invention par une  
40 étape 12 qui, compte tenu de la sub-optimalité de l'algorithme, consiste à corriger les valeurs de pitch  $M_k$  :

En effet, il est possible de trouver à la même itération un pitch  $M_k$  de valeur identique à celle de l'un des pitches déjà calculé. Pour améliorer l'algorithme, l'étape 12 optimise le calcul au-delà du nombre K effectif d'excitations recherché et choisit et/ou regroupe parmi les  $K'(K' < K)$  excitations celles qui donnent le meilleur résultat acoustique. Par exemple, deux excitations dont les valeurs de M sont trop voisines pour être discernées  
45 sont regroupées en une seule excitation. La détermination du niveau de l'excitation apériodique reste quant à elle identique dans les deux modes de réalisation du procédé selon l'invention. Une fois les différentes excitations déterminées par le procédé selon l'invention, l'analyse se poursuit par une quantification selon des procédés connus.

Un mode de réalisation d'un vocodeur permettant la mise en oeuvre du procédé selon l'invention est représenté à la figure 4. Ce dispositif comporte un générateur de bruit 13 délivrant une forme d'onde aléatoire, ou excitation apériodique, K générateurs 14<sub>1</sub> à 14<sub>K</sub> délivrant chacun un train d'ondes périodiques où chaque période du fondamental, "pitch", est notée respectivement  $M_1$  à  $M_K$ . L'excitation apériodique correspond aux sons non voisés comme la plupart des consonnes et les K trains d'ondes périodiques correspondent à des sons voisés comme les voyelles. L'excitation apériodique et les K excitations apériodiques ainsi définies sont affectées respectivement d'un gain  $G_0$  à  $G_K$  modulable dans le temps représenté respectivement par les cercles  
50 15<sub>0</sub> à 15<sub>K</sub>. Les K + 1 excitations sont ensuite injectées simultanément à l'entrée d'un sommateur 16. En sortie du sommateur 16, on obtient les k + 1 excitations superposées qui sont injectées sur une première entrée d'opérande d'un opérateur de multiplication 17. La deuxième entrée d'opérande permet d'ajuster le niveau global

des  $k + 1$  excitations. Après avoir défini le niveau global et la répartition des gains des diverses excitations, tout en respectant la relation suivante  $G_0 + G_1 + \dots + G_k = 1$ , le signal de sortie de l'opérateur 17 est injecté en entrée d'un filtre d'analyse 18, par exemple, un filtre de prédiction d'ordre  $P$  qui en utilisant le procédé d'analyse selon l'invention délivre en sortie un signal de parole synthétique débarrassé des bruits ambiants.

Un procédé de quantification utilisable avec un tel vocodeur est donné à titre d'exemple :

De façon usuelle, le niveau global de l'énergie de la trame est quantifié semi-logarithmiquement sur 5 bits.

En supposant l'existence de  $K = 3$  générateurs d'excitation périodique, et en admettant une quantification non linéaire du pitch sur 6 bits, avec  $M_{\min} = 20$  et  $M_{\max} = 160$ , il faut 18 bits pour définir les trois pitches,  $M_1$ ,  $M_2$  et  $M_3$ .

Enfin, pour s'affranchir des erreurs de transmission affectant le pitch, et ne pas avoir à transmettre le gain  $\beta_0$  du générateur de bruit, on propose plutôt de transmettre les valeurs des niveaux initiaux  $\beta_{i, \text{init}} (i = 1 \dots K)$  des générateurs périodiques de leurs niveaux finaux  $\beta_{i, \text{fin}} (i = 1 \dots K)$  ces niveaux finaux étant atteints sur la durée  $N$  d'une trame, et étant définis par l'expression suivante :

$$(17) \quad \beta_{i, \text{fin}} = \beta_{i, \text{init}} \left( 1 + \alpha_i \frac{N}{M_i} \right)$$

A raison de 2 bits par coefficient, ceci représente un total de  $3(2+2) = 12$  bits, sachant par ailleurs que  $\beta_0$  est le complément à 1 de la somme des  $\beta_i$ .

Le nombre de bits alloués à l'excitation est donc de  $5 + 18 + 12 = 35$ .

En utilisant pour la quantification du filtre de prédiction un procédé similaire à celui mis en oeuvre pour un vocodeur à 1200 bits/seconde, où le débit est d'environ 25 bits par filtre pour une qualité équivalente à celle du 2400 bits/seconde, le débit obtenu est de 2400 bits/seconde au plus pour des trames de 25 ms au moins.

## Revendications

1. Procédé de discrimination de la parole en présence de bruits ambiants pour vocodeur à faible débit du type comportant une excitation périodique, une excitation apériodique et un filtre d'analyse d'ordre  $P$ , caractérisé en ce qu'il consiste pour analyser un signal  $S_n$  composé de la somme d'un nombre  $K$  déterminé d'excitations périodiques et d'une excitation apériodique, à calculer (1) l'autocorrélation globale  $r_m$  du signal  $S_n$ , à calculer (2) les sommes partielles  $t_m$  de l'autocorrélation à court terme  $s_m$  corrélée avec l'autocorrélation globale  $r_m$ , à initialiser (3) un compteur  $k$  et tant que le compteur  $k$  n'atteint pas la valeur maximale  $K$  correspondant au nombre maximal d'excitations périodiques (6), pour chaque incrémentation (4) du compteur  $k$ , après avoir corrigé (7) le calcul des sommes partielles  $t_m$ , à calculer (5) les valeurs du pitch  $M_k$ , du gain  $\beta_k$  et de la pente du gain  $\alpha_k$  de chaque excitation périodique, et à déterminer le niveau de l'excitation apériodique  $\beta_0$  en fonction de l'autocorrélation globale  $r_m$  en début d'analyse et de l'autocorrélation  $r_m$  en fin d'analyse.
2. Procédé selon la revendication 1, caractérisé en ce que l'autocorrélation globale  $r_m$  est bornée entre une valeur minimale  $m = -P$ , où  $P$  correspond à l'ordre de prédiction à court terme du filtre d'analyse, et une valeur maximale  $m = M_{\max} + P$ , où  $M_{\max}$  correspond à la valeur maximale du pitch  $M_k$  considéré.
3. Procédé selon l'une quelconque des revendications 1 et 2, caractérisé en ce que les valeurs  $M_k$ ,  $\beta_k$  et  $\alpha_k$  sont déterminées en minimisant par une méthode des moindres carrés la distance  $d$  définie comme étant la distance entre le vecteur  $R$  de l'autocorrélation globale  $r_m$  et le vecteur  $Sl_k$  de l'autocorrélation à court terme  $s_m$  convolué avec la même excitation périodique, multiplié par le gain  $\beta_k$ .
4. Procédé selon l'une quelconque des revendications 1 à 3, caractérisé en ce que les sommes partielles  $t_m$  sont modifiées en soustrayant des sommes partielles  $t_m$ , les valeurs de l'autocorrélation d'échantillons  $s_m$  de l'autocorrélation à court terme et en remplaçant les anciennes sommes partielles  $t_m$  par les nouvelles sommes  $t_m'$  résultant de la soustraction.
5. Procédé selon l'une quelconque des revendications 1 à 4, caractérisé en ce qu'il consiste en outre à pré-traiter (8) le signal d'entrée  $S_n$  pour affiner le calcul (1) de l'autocorrélation globale  $r_m$ , à pondérer (9) l'au-

- 5      tocorrélation globale  $r_m$  par une fenêtre de pondération limitant le nombre de répliques de l'autocorrélation, à tester (10) après le calcul (5) des valeurs de  $M_k$ ,  $\beta_k$  et  $\alpha_k$ , si la valeur du compteur  $k$  est strictement inférieure à deux, alors à incrémenter (4) le compteur  $k$ , sinon à recalculer (10) les valeurs  $\alpha_i$  et  $\beta_i$  respectivement à  $\alpha_k$  et  $\beta_k$  en conservant les valeurs de  $M_k$  précédemment calculées, de  $i$  égal à un, à  $i$  égal à la valeur du compteur  $k$ , et tant que la valeur du compteur  $k$  n'a pas atteint la valeur maximale  $K$ , à incrémenter (4) le compteur  $k$ , et quand la valeur maximale  $K$  est atteinte, à regrouper en une seule excitation, deux excitations dont les valeurs  $M_k$  sont trop voisines pour être discernables.
- 10      6. Procédé selon la revendication 5, caractérisé en ce que le prétraitement (8) du signal d'entrée  $S_n$  est une prédiction auto-adaptative permettant un blanchissement du spectre du signal d'entrée  $S_n$ .
- 15      7. Procédé selon la revendication 6, caractérisé en ce que la fenêtre de pondération de l'autocorrélation globale  $r_m$  est une fonction non croissante du temps dont la largeur est choisie supérieure à l'intervalle maximal d'analyse.
- 20      8. Vocodeur pour la mise en oeuvre du procédé selon l'une quelconque des revendications 1 à 7, caractérisé en ce qu'il comporte un générateur de bruit (13) délivrant une excitation apériodique,  $K$  générateurs (14<sub>1</sub> à 14<sub>K</sub>) délivrant  $K$  excitations périodiques, les  $K + 1$  excitations étant affectées respectivement d'un gain ( $G_0$  à  $G_K$ ) modulable dans le temps et tel que la somme des gains ( $G_0$  à  $G_K$ ) est égale à un, un circuit sommateur (16) recevant les  $K + 1$  excitations dont la sortie est couplée à une première entrée d'opérande d'un opérateur de multiplication (17), la deuxième entrée d'opérande permettant d'ajuster le niveau global des  $K + 1$  excitations, et un filtre d'analyse (18) assurant l'analyse des  $k + 1$  excitations pour en déduire le signal synthétique représentatif de la parole en présence de bruits ambiants.
- 25      9. Vocodeur selon la revendication 8, caractérisé en ce que le filtre d'analyse (18) est un filtre de prédiction linéaire d'ordre  $P = 10$ .

30

35

40

45

50

55

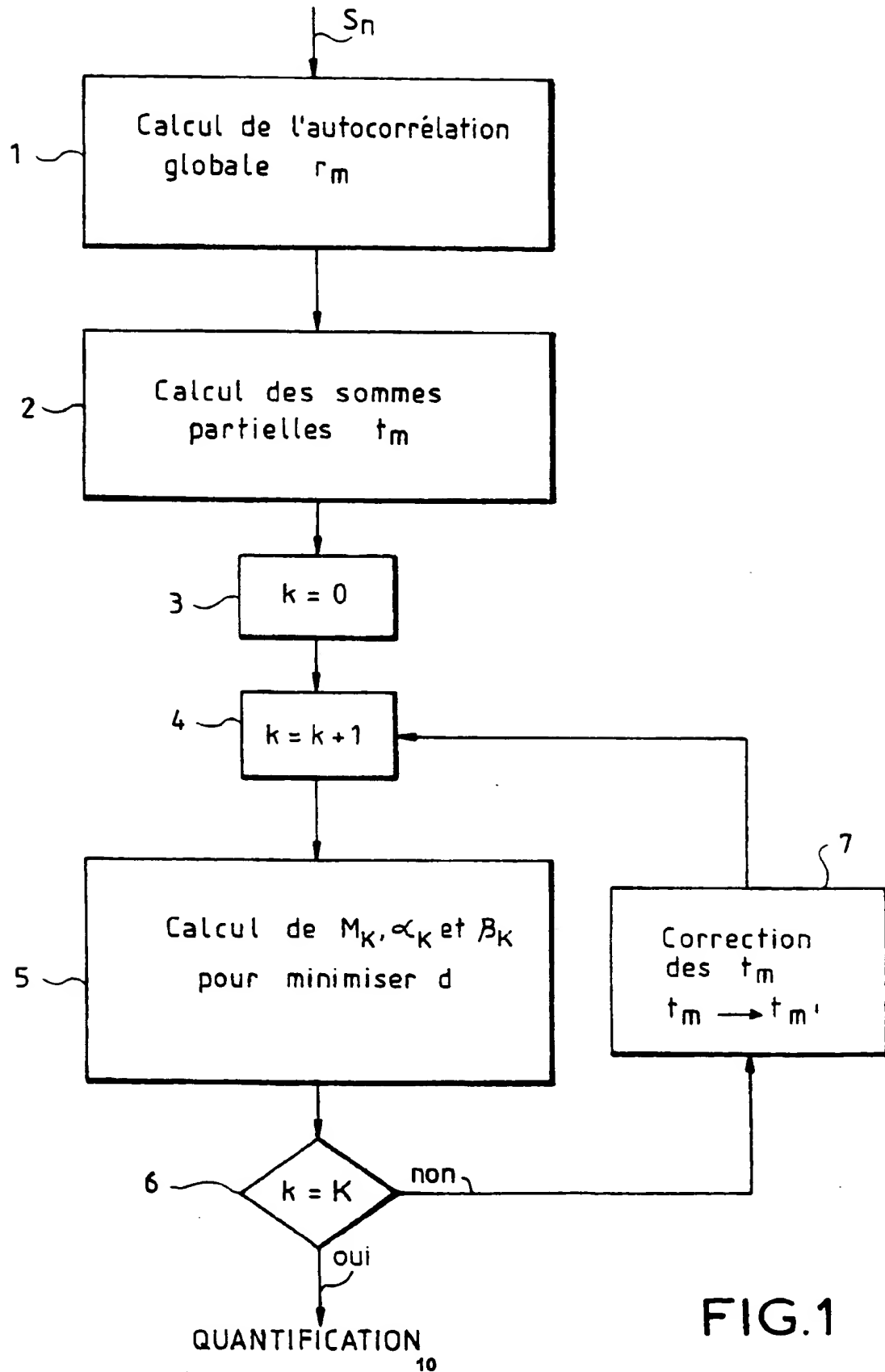


FIG.1

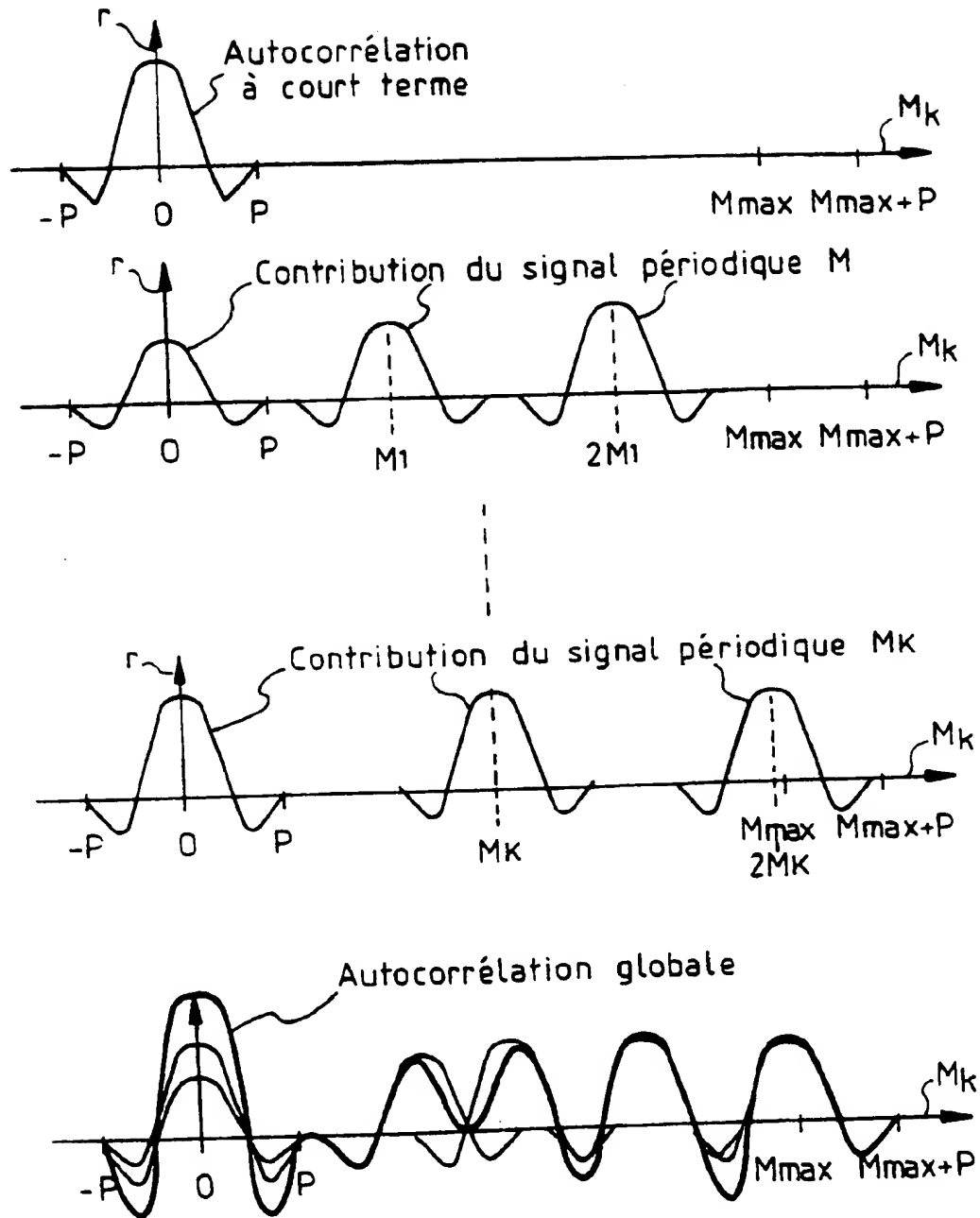


FIG.2

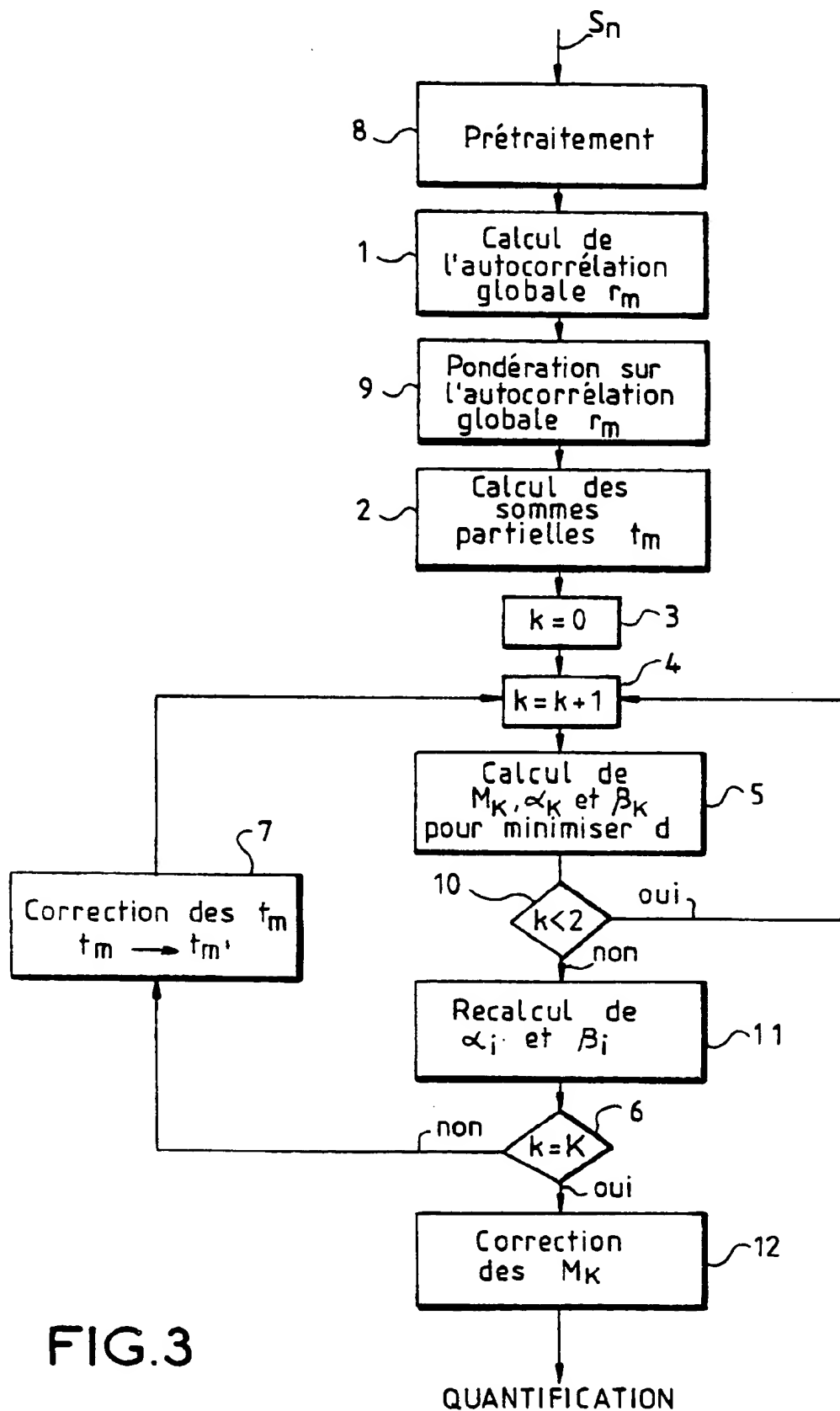


FIG.3

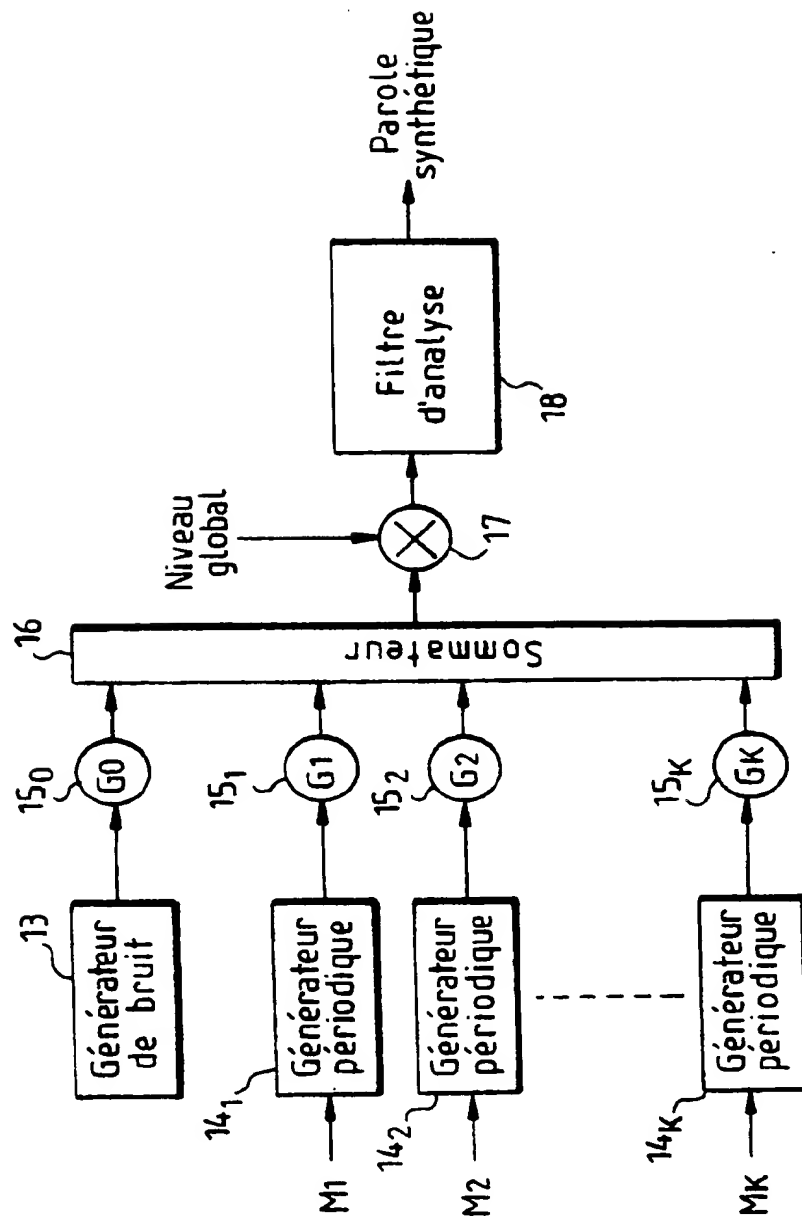


FIG. 4

Office européen  
des brevets

## RAPPORT DE RECHERCHE EUROPEENNE

Numero de la demande

EP 93 40 2670

| DOCUMENTS CONSIDERES COMME PERTINENTS   |   |   |  |
|---|---|---|--|
| Catégorie   | Citation du document avec indication, en cas de besoin, des parties pertinentes   | Revendication concernée   | CLASSEMENT DE LA DEMANDE (Int.Cl.5)                  |
| 8 A   | NEC RESEARCH AND DEVELOPMENT<br>no. 85 , Avril 1987 , TOKYO JP<br>pages 112 - 121<br>ISHIKAWA IKEDA 'Narrow to medium bands<br>speech codec family based on LPC<br>technique'<br>* pages 118-119, section 5, figure 7 *   | 1,3   | G10L9/14<br>G10L9/08                                 |
| 8 A   | INTERNATIONAL CONFERENCE ON ACOUSTICS<br>SPEECH AND SIGNAL PROCESSING<br>vol. 2 , 7 Avril 1986 , TOKYO JAPAN<br>pages 1261 - 1264<br>SAGAYAMA, ITAKURA 'Duality theory of<br>composite sinusoidal modeling and linear<br>prediction'<br>* pages 1261-1262, section II * | 1   |  |
| 11 A  | INTERNATIONAL CONFERENCE ON ACOUSTICS<br>SPEECH AND SIGNAL PROCESSING<br>vol. 1 , 6 Avril 1987 , DALLAS TEXAS<br>pages 356 - 359<br>JAIN, XU 'Autocorrelation distortion<br>function for improved AR modeling'<br>* pages 357-358, section II *                         | 1,3   |  |
| 6 A   | IEEE GLOBAL TELECOMMUNICATION CONFERENCE<br>vol. 1 , 28 Novembre 1988 , FLORIDA USA<br>pages 285 - 289<br>SUKKAR ET AL 'LPC excitation based on zinc<br>function decomposition'<br>* le document en entier *  | 1   |  |
| Le présent rapport a été établi pour toutes les revendications  |   |   | DOMAINES TECHNIQUES<br>RECHERCHES (Int.Cl.5)<br>G10L |
| Lieu de la recherche<br><b>LA HAYE</b>  |   | Date d'achèvement de la recherche<br><b>21 Janvier 1994</b>   | Examinateur<br><b>Farassopoulos, A</b>               |
| CATEGORIE DES DOCUMENTS CITES   |   | T : théorie ou principe à la base de l'invention<br>E : document de brevet antérieur, mais publié à la date de dépôt ou après cette date<br>D : cité dans la demande<br>L : cité pour d'autres raisons<br>& : membre de la même famille, document correspondant |  |
| X : particulièrement pertinent à lui seul<br>Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie<br>A : arrière-plan technologique<br>O : divulgation non-écrite<br>P : document intercalaire |   |   |  |

EPO FORM 150 (3.82) (P4/C2)





Office européen  
des brevets

# RAPPORT DE RECHERCHE EUROPEENNE

Numero de la demande  
EP 93 40 2670

| DOCUMENTS CONSIDERES COMME PERTINENTS   |  |  |   |
|---|--|--|---|
| Catégorie   | Citation du document avec indication, en cas de besoin, des parties pertinentes  | Revendication concernée                              | CLASSEMENT DE LA DEMANDE (Int.Cl.5)       |
| A   | <p>INTERNATIONAL CONFERENCE ON ACOUSTICS<br/>SPEECH AND SIGNAL PROCESSING<br/>vol. 2 , 6 Avril 1987 , DALLAS TEXAS<br/>pages 948 - 951<br/>LIENARD 'Speech analysis and<br/>reconstruction using short time,<br/>elementary waveforms'<br/>* page 948, " abstract ", section I; page<br/>949, section III *</p> <p>-----</p> | 1  |   |
|   |  |  | DOMAINES TECHNIQUES RECHERCHES (Int.Cl.5) |
| Le présent rapport a été établi pour toutes les revendications  |  |  |   |
| Lieu de la recherche<br>LA HAYE   |  | Date d'achèvement de la recherche<br>21 Janvier 1994 | Examinateur<br>Farassopoulos, A           |
| <p><b>CATEGORIE DES DOCUMENTS CITES</b></p> <p>X : particulièrement pertinent à lui seul<br/>Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie<br/>A : arrière-plan technologique<br/>O : divulgation non-écrite<br/>P : document intercalaire</p> <p>T : théorie ou principe à la base de l'invention<br/>E : document de brevet antérieur, mais publié à la date de dépôt ou après cette date<br/>D : cité dans la demande<br/>L : cité pour d'autres raisons<br/>Δ : membre de la même famille, document correspondant</p> |  |  |   |

EPO FORM 1503 03.92 (P04/C02)

**This Page Blank (uspto)**